March 2020

Using EEG to Predict Search Intent and to Control IoT Devices

N/A

Using EEG to Predict Search Intent and to Control IoT Devices

ABSTRACT

It has been demonstrated that electroencephalogram (EEG) readings can be mapped to images, audio, or language perceived by a user, such that a machine-learning model trained on EEG readings can infer the user's thoughts. This disclosure describes techniques that use a head-mounted device, e.g., smart glasses, to measure, with user permission, a user's EEG, and based on the measurement, infer the user's intent or reconstruct the user's thoughts, e.g., images or audio therein. In situations where a user can only partially recall a search phrase, the techniques enable determination of the user's search intent from the user's EEG. In other situations, the techniques enable the user to control Internet-of-Things (IoT) sensors, e.g., the user can change the temperature of a room by visualizing a thermostat.

KEYWORDS

- Electroencephalogram (EEG)
- Search intent
- Brain-computer interface
- Brain-machine interface
- Internet-of-Things (IoT)
- Machine learning

BACKGROUND

It has been demonstrated that bodily electromagnetic emissions (EM), e.g., fMRI, electroencephalogram (EEG) readings, IR, pulse, etc. can be mapped to images, audio, or language perceived by a user, such that a machine-learning model trained on such EM emissions can infer the user's thoughts. Although the fMRI signal more reliably produces an estimate of the user's intent, the MRI machine itself is not portable. On the other hand, EEG readings can be obtained relatively easily with portable, wearable head-mounted devices, e.g., smart glasses. Until recently, EEG as a signal was relatively noisy and had some reliability issues in determining user attention. Recent advances have shown the EEG signal to be a relatively accurate determiner of user attention and intent.

When users search for information using text or images, their intent can be ambiguous due to the noisy channel between the user and the search interface, e.g., the keyboard, the imprecise language, etc. Besides the difficulty to estimate exactly the user's intent, it is often the case that the user is unable to properly define, understand, or communicate their intent.

DESCRIPTION

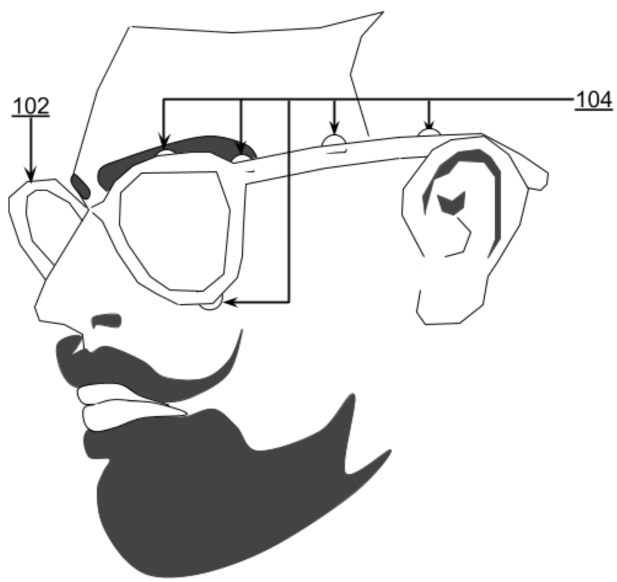

Fig. 1: Using a head-mounted device to measure EEG

As illustrated in Fig. 1, this disclosure describes techniques that use a head-mounted device (102), e.g., smart glasses, to measure non-invasively and with user permission, a user's EEG using sensors (104). The measurements are used to infer the user's intent or reconstruct the user's thoughts, e.g., images or audio therein. This extracted visual information can then be used to improve the understanding of user intent to refine search results and to improve the relevance of user queries. Effectively, the body's ambient emissions reduce or eliminate the aforesaid noisy channel between the user and the search interface, and help recover of the user's original search intent.

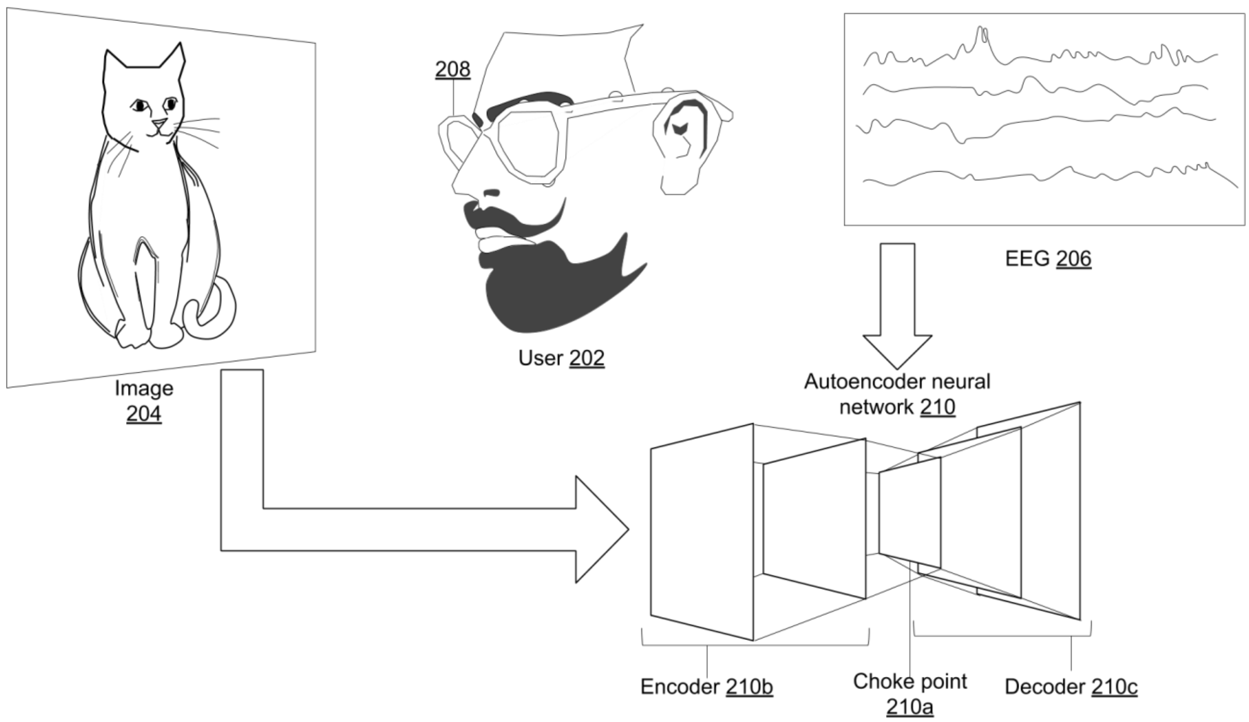

Fig. 2: Training a machine-learning model to recognize user intent

Fig. 2 illustrates training of a machine-learning model to recognize user intent. A user (202) wearing a head-mounted device (208) capable of measuring EEG watches a series of images (204). For each image, the head-mounted device captures the user's EEG (206). The image and the EEG are fed to a machine-learning model, e.g., an autoencoder neural network (210).

The machine-learning model is trained to map the EEG to the image. Once trained, the machine-learning model can infer or reconstruct an image in the user's thoughts based on EEG readings. Effectively, EEG signals are mapped to the latent embedding space of an autoencoder deep learning neural network, trained on pairs of *<EEG, image>*.

As illustrated in Fig. 2, the stages of an autoencoder neural network decline in size until they reach a certain size known as choke point (210a), following which they re-expand. The declining stages comprise an encoder (210b), and the expanding stages comprise a decoder (210c). Images are fed into the topmost layer of the encoder, and the decreasing size of the deeper layers causes the autoencoder to learn a latent, compressed, generalized representation or encoding of the images being fed to it.

The choke point at the center of the autoencoder is the encoded vector or embedding representing the images. Encoding is done using convolutions, as in a convolutional neural network (CNN). The decoder component of the autoencoder reconstructs the image from the latent embedding/choke point using deconvolution.

During deconvolution, the network upscales and reconstructs the images in a reverse fashion. During training (or error optimization) procedures such as forward/backward propagation, the weights of the network are updated based on, e.g., gradient descent, until the images reproduced by the convolution-deconvolution procedure are of sufficient quality to the extent the chosen network architecture can restore or regenerate them.

The training of the neural network enables the generation of an image from an embedding of the image, e.g., by using just the decoder. The latent embedding space of the neural network is mapped to the set of EEG signals given off by the user. Once mapped, an EEG signal-set can be used to reconstruct an image. The aforesaid mapping is done by training a second neural network whose cost function is a paired loss jointly measuring the following errors:

- the error between the EEG signal-set and the predicted image embedding in the latent space of the second neural network, and
- the error from the latent image embedding space and the space the second neural network is learning.

This causes the second network to learn an efficient mapping between the two spaces. To train the neural network, the image seen by the user is synchronized in time with its corresponding EEG signal.

In a similar manner, the machine-learning model can be trained on video, audio, text, etc. For example, for video, EEG signals are mapped to the latent embedding space of an autoencoder, trained on pairs of *<EEG, frame>* sequences of video.

The machine-learning model that maps EEG signals to images, videos, audio, or text can be personalized to a user, e.g., by having the user train their own personalized model. The machine-learning model can also be trained using larger crowdsourced models and transfer-learning techniques such as distillation training. Under transfer learning, a model trained on a large population of users can be refined by forcing the EEG-to-latent image space embeddings from the global/crowdsourced database to a space closer to the user's EEG signals. This results in a smaller training time since the user does not have to train the model from scratch and can provide improved precision/accuracy performance.

Some example applications are as follows:

Example 1: Searching for a video. A user is searching a video-sharing website for a video of their favorite sporting event last watched years ago. The user does not remember key details such as the names of the teams playing, the location the game was played, etc. The user does, however, have memories of sequences of frames from the event as a visual, and memories of the sounds of the announcers of the event. The user's EEG signals are mapped to the images and sounds in the user's thoughts, and a search-intent dialog box opens on the user's computing device with those images and sounds. If the user confirms the images and sounds in the search-intent box, the corresponding video is played.

Example 2: Searching for an image. A user is conducting an image-search for a certain kind of animal that they saw in an advertisement. The user believes the name of the animal is cheetah and writes "cheetah" in the search box, when in fact the animal is a snow leopard. The search fails to bring up results satisfactory to the user. Per the techniques described herein, latent visual information from the user's thoughts, obtained via processing of the user's EEG by a machine-learning model, improves the image search by determining search intent. For example, clickable pictures of both a cheetah and a snow leopard are provided with a prompt "Did you mean snow leopard?" Similarly, users can search for products, wallpapers, etc., whose images they remember but whose names they do not.

Example 3: Deals appropriate to the user. A user is watching their favorite program on television, mobile device, or browser while wearing an EEG-equipped head-mounted device. An advertisement pops up. The advertisement that the user is viewing is recognized by the machine learning model that processes the user's EEG. A coupon or deal appropriate to the user is brought up for the user to look at, if the user has opted to receive such deals. Alternatively, such information is embedded into their search results the next time the user conducts a search related to the content of the ad. Even if the user does not remember the name of the advertisement or the company or product being advertised, using the context that the user previously saw the ad can help boost search results, e.g., bring more appropriate search results to the user. The content can be used to rerank search results, e.g., by boosting the products of the advertiser or company, or the overall product category. The suggestion-phrase “did you mean <name-of-advertisement?>” can be added to enable the user to refine the search. This is especially useful if the user is getting close to this query but not quite the whole way.

In the case of pairing a search result improvement or refinement with video or images the user is currently looking at, the search signals can be deduced directly from video or audio sensors on the user’s devices. However, many users do not grant permission for audio and video sensors to continuously gather data. Using EEG signals, as described herein, provides a private and personalized way to determine search intent. The use of EEG also ensures that the search intent came from a particular user and not from someone nearby in the same room. The user can disable provision of EEG information by simply removing the wearable device or by turning off the feature.

The described techniques to infer user intent from EEG signals can also be used to construct neural (brain-computer) interfaces, e.g., to control IoT devices. For example, the user can train the machine-learner to recognize that a specific image thought out in the user’s mind corresponds to a command, e.g., “close the blinds,” “turn off the lights,” etc.

An example calibration procedure for neural control of IoT devices is as follows.

1. The user chooses an image from a set of command images (or provides their own);
2. The user wears an EEG-equipped head-mounted device;
3. The user looks at a black image to prevent accidental leakage of information from something just looked at; and
4. The user looks at the training image while the user’s EEG is captured (e.g., after a button is pressed) for the purpose of training the machine-learning model. A black image is presented between consecutive command images.

After training, the user can now control IoT devices by thinking of the corresponding image. For example, the thought of a red thermostat can trigger an increase of room temperature, while the thought of a blue thermostat can trigger a decrease of room temperature.

A feedback signal as to whether the thought was correctly decoded can be given by the user tactically, verbally, or by thinking. For example, the thought of a red cross (❌)can be mapped to a negative feedback signal while the thought of green check-mark (✔)can be mapped to a positive feedback signal. In this manner, the machine-learning model can refine its accuracy or understanding of the correct EEG signal. The feature of thought-based training feedback is especially useful when the user's thoughts wander. A confirmation step, e.g., an opportunity for verbal or tactical cancelation, can be added as a verification phase, in case the thought-based feedback signal becomes noisy.

Alternative to using images for the purpose of commanding IoT sensors, the user can also train the machine-learning model to accept commands based on inaudibly thought-out phrases. For example, the user can think about the phrase “close the blinds,” for the machine-learning model to determine user intent and close the blinds.

Additional examples of applications for EEG-based commands or search refinements include the following:

- improving navigation search results, e.g., the user vaguely remembers a particular route but cannot verbally articulate it, and EEG-based intent determination refines routing such that the user stays in familiar territory while traveling;
- sending text messages without touching or typing on a device;
- unlocking car doors without touching a key;

- sending emojis by thinking of an emoji picture, e.g., the user thinks of cats and triggers a messaging application to select and send from a dataset of appropriate images a cute picture of a cat or an image most similar to the picture the user is thinking of;
- unlocking electronic devices by thinking of a passphrase or a particular image; etc.

Still further, the described techniques can be used to improve the accuracy of automatic machine translation, e.g., provided via a smartphone and/or wearable device by obtaining EEG from both the speaker and listener in a conversation. The translation can be based on intent (as determined from EEG) in addition to spoken phrases which can provide improved accuracy over translation that is based solely on spoken phrases. Such translation can address the loss of semantic meaning that occurs when translation models are trained only on spoken phrases. The training of machine translation models can be improved by using the intent of the phrases being translated, as determined from EEG.

The quality of automatic speech recognition can also be improved by the use of EEG. When the user utters a phrase to a device, the quality of speech recognition performed by the device can be improved by providing additional input from the user’s EEG which is indicative of the user’s intent. Still further, silent communication between two humans can also be enabled by use of respective EEG.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user’s social network, social actions or activities, profession, a user’s preferences, or a user’s current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used so that personally identifiable information is removed.

For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques that use a head-mounted device, e.g., smart glasses, to measure, with user permission, a user's EEG, and based on the measurement, infer the user's intent or reconstruct the user's thoughts, e.g., images or audio therein. In situations where a user can only partially recall a search phrase, the techniques enable determination of the user's search intent from the user's EEG. In other situations, the techniques enable the user to control Internet-of-Things (IoT) sensors, e.g., the user can change the temperature of a room by visualizing a thermostat.

REFERENCES

[1] Grigory Rashkov, Anatoly Bobe, Dmitry Fastovets, and Maria Komarova, "Natural image reconstruction from brain waves: a novel visual BCI system with native feedback," https://www.biorxiv.org/content/10.1101/787101v2.full accessed on Feb. 29, 2020.

[2] "Neural network reconstructs human thoughts from brain waves in real time," https://techxplore.com/news/2019-10-neural-network-reconstructs-human-thoughts.html accessed on Feb. 29, 2020.

[3] "Electroencephalography," https://en.wikipedia.org/wiki/Electroencephalography accessed on Feb. 29, 2020.